CA/EM (Q12L13) 50% (10/20)

✗ 1. Which of these algorithms can be used to fill the missing values

- (A) KNN for regression
- (B) KNN for classification
- (C) both
- (D) I do not know

✗ 2. Bagging is a technique used to reduce

- (A) the variance of our predictions
- (B) the bias of our predictions
- (C) both
- (D) I do not know

✗ 3. How can Ensemble methods be constructed?

- (A) By manipulating the training set
- (B) By manipulating the input features
- (C) By manipulating the class labels
- (D) By manipulating the learning algorithm
- (E) All of them
- (F) None
- (G) I do not know

✓ 4. Repeatedly sampling observations are taken

- (A) from general population
- (B) original sample data set
- (C) I do not know
- (D) None

✓ 5. Random Forest differs from bagging

- (A) by a random sample of m predictors
- (B) by bootstrapped training samples
- (C) by adaptive sampling
- (D) I do not know

✗ 6. Boosting differs from bagging

- (A) by a random sample of m predictors
- (B) by bootstrapped training samples
- (C) by adaptive sampling
- (D) I do not know

✗ 7. Averaging many highly correlated quantities

- (A) lead to as large of a reduction in variance
- (B) does not lead to as large of a reduction in variance
- (C) lead to as large of a reduction in bias
- (D) I do not know

✗ 8. We can perform a Random forest in R using the function

- (A) randomForest()
- (B) rf()
- (C) randomF()
- (D) boot()
- (E) I do not know

✗ 9. Random Forest works

- (A) for classification
- (B) for regression
- (C) both
- (D) I do not know

✗ 10. Cluster Analysis is

- (A) Unsupervised learning technique
- (B) Supervised learning technique
- (C) I do not know

✓ **11. Distance between records and distance between clusters are the same**

- (A) True
- (B) False
- (C) I do not know

✓ **12. Which of these is the measure of between clusters distance?**

- (A) Single link
- (B) Complete link
- (C) Average link
- (D) Centroid
- (E) All of them
- (F) I do not know

✓ **13. Single link is**

- (A) the smallest distance between an element in one cluster and an element in the other
- (B) the largest distance between an element in one cluster and an element in the other
- (C) the average distance between an element in one cluster and an element in the other
- (D) distance between the centroids of two clusters
- (E) I do not know

✓ **14. Complete link is**

- (A) the smallest distance between an element in one cluster and an element in the other
- (B) the largest distance between an element in one cluster and an element in the other
- (C) the average distance between an element in one cluster and an element in the other
- (D) distance between the centroids of two clusters
- (E) I do not know

✓ **15. Which of these is the nested algorithm of clustering?**

- (A) Hierarchical clustering
- (B) k-means
- (C) Knn
- (D) I do not know

✗ 16. Which of these is the unnested algorithm of clustering?

- A Hierarchical clustering
- B k-means
- C Knn
- D I do not know

✗ 17. Which of these is the type of hierarchical clustering?

- A Agglomerative Methods
- B Divisive Methods
- C Both
- D I do not know

✓ 18. This function can be used to perform hierarchical clustering in R

- A hclust()
- B cluster()
- C hierarchical ()
- D I do not know

✓ 19. This function can be used to perform k-means clustering in R

- A kmeans()
- B kclust()
- C kmenscl()
- D I do not know

✓ 20. Do we need to worry about scaling in clustering?

- A Yes
- B No
- C I do not know